Supplementary Information

Supplementary Materials and Methods

Patient sample processing

Samples were sent to the coordinating center (OHSU; IRB 4422; NCT01728402), where they were coded and processed. Specific names of centers associated with each specimen were coded (code 1-14) and centers providing less than two samples were aggregated together and given one center identifier (code 15). Mononuclear cells were isolated by Ficoll gradient centrifugation and/or red cell lysis from freshly obtained bone marrow aspirates or peripheral blood draws. Cell pellets were snap-frozen in liquid nitrogen for subsequent DNA isolation (Qiagen, DNeasy Blood & Tissue Kit), freshly pelleted cells were lysed immediately in guanidinium thiocyanate (GTC) lysate for subsequent RNA isolation (Qiagen, RNeasy Mini Kit). All samples were collected between 2001-2015.

Clinical, prognostic, genetic, cytogenetic and pathologic laboratory values, as well as treatment and outcome data, were manually curated from the electronic medical records of the patient. Patients were assigned a specific diagnosis in accordance with WHO 2017 criteria. Cases fulfilling 2017 WHO classification criteria for RARS-T, MDS or MPN were excluded. Cases with *BCR/ABL1, PDGFRA, PDGFRB,* or *FGFR1* rearrangement were also excluded. Totally, we have collected clinical information of 41 CNL, 28 aCML, 14 MDS/MPN-U, 13 MPN-U, 30 CMML and 57 patients with NA or ambiguous diagnosis. Notably, alternative diagnosis indicates cases with transformed AML, mastocytosis, reactive neutrophilia or MPN diagnosis, which are therefore excluded from the analysis.

WES processing

Initial data processing and alignments were performed with in-house workflows. BWA MEM[1] was used to align the read pairs for each sample-lane FASTQ file. As part of this process, the flowcell and lane information was kept as part of the read group of the resulting SAM file. The Genome Analysis Toolkit (v3.3)[2] and the bundled Picard were used for alignment post-processing. The files contained within the Broad's bundle 2.8 were used including their version of the build 37 human genome (These files were downloaded from: ftp://ftp.broadinstitute.org/bundle/2.8/b37/). The following steps were performed per sample-lane SAM file generated for each CaptureGroup:

1. The SAM files were sorted and converted to BAM via SortSam
2. MarkDuplicates was run, marking both lane level standard and optical duplicates

3. The reads were realigned around indels from the reads RealignerTargetCreator/IndelRealigner.
4. Base Quality Score Recalibration

The resulting BAM files were then aggregated by the sample and an additional round of MarkDuplicates. Indel realignment was carried out again across the cohort of samples. Genotyping was performed using the UnifiedGenotyper tool that is part of GATK. These variants were assigned to their most deleterious effect on Ensembl transcripts using Ensembl VEP v83 on GRCh37 and further curated. CALR indels were called from Pindel[3].

Variant Calling

Variant calling was similar to a previous study[4]. Briefly, since no paired normal tissue controls were available, we compiled a list of genes associated with human hematologic malignancies according to these two papers[5,6]. In total, 170 genes were selected (Supplementary Table 1). The following filters were used: 1) excluding variants at a frequency greater than 0.1% in the ExAC database, and excluding variants at a frequency greater than 20% in BeatAML normal controls[7]; 2) including variant types: Missense; Frameshift; Stop gain/loss; Inframe insertion/deletion; Protein altering; and Tandem duplication for 132 genes listed in Supplementary table 1 (regular black font). 4) In addition, only frameshift, stop gain/loss and Inframe insertion/deletion variants are considered for the 38 genes in bold red font in Supplementary Table 1. 5) Variants were further manually curated, excluding variants seen in dbSNP, but not in Cosmic; predicted 'tolerated' by sift and predicted 'benign' by polyphen; some TCGA and Jaiswal variants were added back based on convincing VAF pattern and known pathogenic role. 6) For the final list, only variants in genes highly relevant to hematology malignancies from the knowledge of AML literature were included, and only inframe indels *of CALR* were included.

Sanger sequencing

Sanger sequencing was performed as previously described[8]. *ASXL1* exons were amplified using forward primer 5'-GCAATTTAGGTATGAAAGCCAGC-3' and reverse primer 5'-CTTTCAGCATTTTGACGGCAACC-3'. PCR products were purified using Amicon Ultra Centrifugal Filters (#UFC503096, Millipore) and sequenced with the same primers by Eurofins operons.

RNAseq Expression processing

The Subjunc aligner [9] was used to align reads to the GRCh37 version of the human genome. These alignments were summarized at the gene-level relative to Ensembl 83 gene models using featureCounts[9]. RNA genotyping was performed using the same protocol as the WES.

Copy number variations

Copy number variations were called using CNVkit [10]. Two reference normalization approaches were carried out depending on the library used. For samples run using the Nimblegen library, we utilized three available skin controls from another project. For samples run using the Nextera library prep, we utilized 50 normal samples from the BeatAML project[7]. Segmentation was performed using DNACopy[11] and the segmented data was summarized per sample and region using CNTools (http://bioconductor.org/packages/release/bioc/html/CNTools.html).

Fusions

Fusions have been generated using STAR-Fusion v1.3.2[12] and Tophat v2.0.14[13]. As fusion calling is necessarily based on the gene models and other annotations utilized (and provided) by a given fusion detection method, we first annotated all the Tophat-fusion calls relative to the STAR-fusion resource gene models to facilitate comparison[12]. We then summarized the fusions treating the left and right genes interchangeably (i.e. *BCR-ABL1* and *ABL1-BCR* would be considered the same fusion) and compared the fusion calls between the callers. High relevant fusion is defined by fusions that are detected by both algorithms, not seen in normal controls and with a fusion allele frequency (FAF) higher than 10%. For the final variant list, only two known pathogenic fusions (*FLT3-MYO18A* and *ABL1-ETV6*) were included[14].

Clustering of the patient samples

We used the Consensus Clustering approach[15] which provides robust clusters based on the expression data by repeatedly clustering random subsets (genes and sample) of the data and recording whether samples cluster together at each repetition for a given number of clusters (k). For this analysis, we used the 80 patient samples and the top 2,000 most variable genes. We clustered them using hierarchical clustering based on the magnitude of their expression levels (Euclidean distance) repeating the procedure 10,000 times. We chose k=7 due to consideration of the cluster definitions as well as comparison with simulated null distributions. We do note, however, that none of the k clusters was strongly defined or performed substantially better than the others.

Supplementary Table 1. List of queried hematopoietic genes.

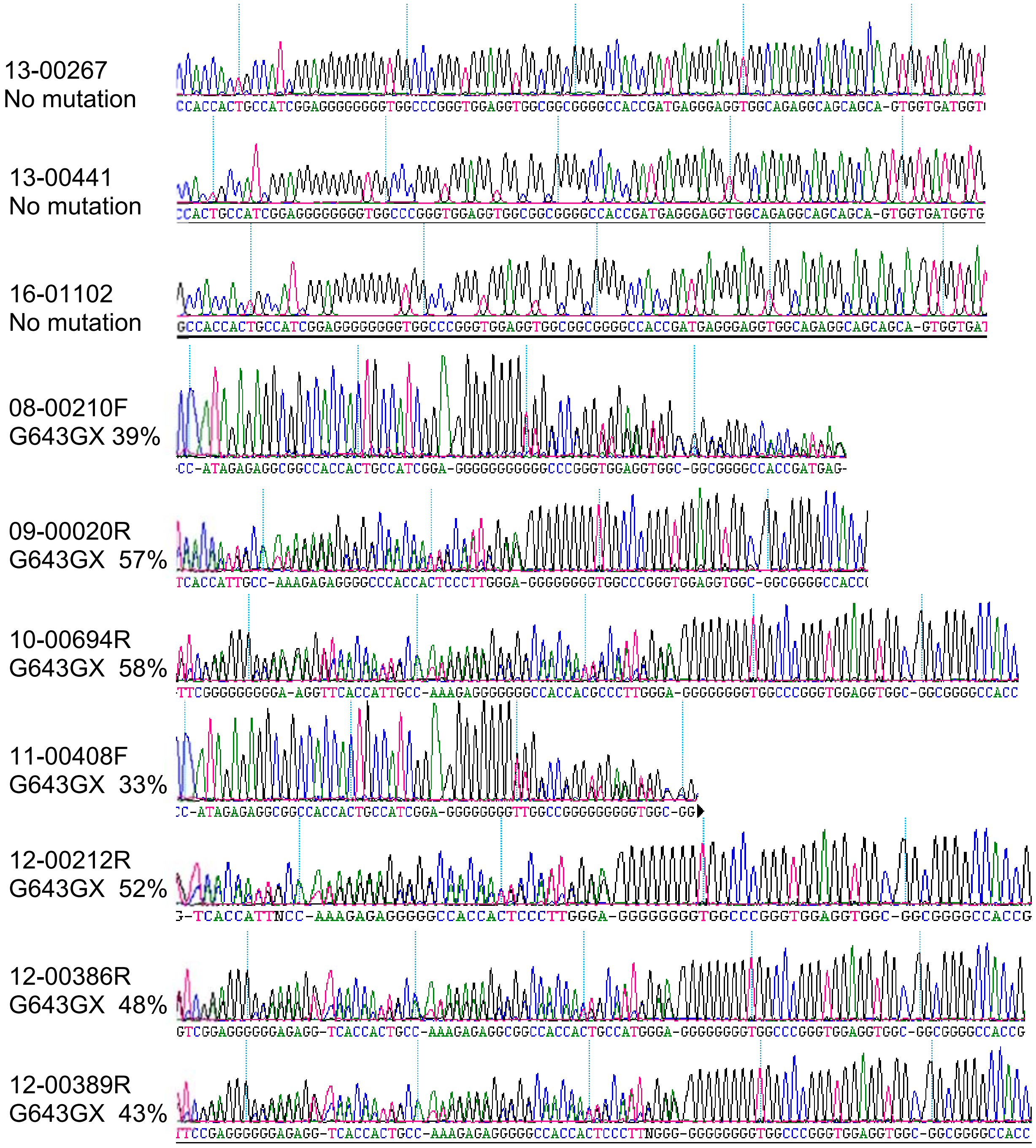
ABL1	*ARID1A*	***ASXL1***	***ASXL2***	*BCL10*	*BCL11B*	*BCL6*	***BCOR***	***BCORL1***	***BIRC3***
BRAF	***BRCC3***	*BTG1*	*BTG2*	*CARD11*	*CBFB*	*CBL*	*CBLB*	*CCND2*	*CCND3*
CD58	*CD70*	*CD79A*	*CD79B*	*CDKN2A*	*CDKN2B*	*CEBPA*	*CHD2*	*CNOT3*	*CREBBP*
CRLF2	*CSF1R*	*CSF3R*	*CTCF*	***CUX1***	*DDX3X*	*DIS3*	*DNMT3A*	*EBF1*	*EED*
EP300	*ETNK1*	*ETV6*	*EZH2*	*EZR*	***FAM46C***	***FAS***	***FBXO11***	*FBXW7*	*FLT3*
FOXP1	*FYN*	*GATA1*	*GATA2*	*GATA3*	*GNA13*	*GNAS*	*GNB1*	*HIST1H1B*	*HIST1H1C*
HIST1H1D	*HIST1H1E*	*HIST1H3B*	*HLA-A*	*ID3*	*IDH1*	*IDH2*	*IKBKB*	***IKZF1***	***IKZF2***
IKZF3	*IL7R*	*INTS12*	*IRF4*	*IRF8*	*JAK1*	*JAK2*	*JAK3*	***JARID2***	***KDM6A***
KIT	*KLHL6*	***KMT2A***	***KMT2C***	***KMT2D***	*KRAS*	***LEF1***	*LRRK2*	***LTB***	***LUC7L2***
MALT1	*MAP2K1*	*MAP3K14*	*MED12*	*MEF2B*	*MPL*	*MXRA5*	*MYD88*	*NF1*	***NFE2***
NOTCH1	***NOTCH2***	*NPM1*	*NRAS*	*NTRK2*	*NTRK3*	*P2RY8*	***PAPD5***	***PAX5***	***PDS5B***
PDSS2	*PHF6*	*PIK3CA*	***POT1***	*POU2AF1*	*POU2F2*	***PPM1D***	*PRDM1*	*PRPF40B*	*PRPF8*
PTEN	*PTPN1*	*PTPN11*	***RAD21***	***RAD21L1***	*RBBP4*	*RHOA*	*RIT1*	*RPL10*	***RPL5***
RPS15	*RPS2*	*RUNX1*	*SETBP1*	*SF3A1*	*SF3B1*	***SGK1***	***SH2B3***	*SMC1A*	*SMC3*
SOCS1	*SPRY4*	*SRSF2*	***STAG1***	***STAG2***	*STAT3*	*STAT5A*	*STAT5B*	*STAT6*	***SUZ12***
SWAP70	***TBL1XR1***	*TCF3*	***TET1***	*TET2*	***TMEM30A***	*TNF*	*TNFAIP3*	*TNFRSF14*	*TP53*
TRAF3	*TYW1*	*U2AF1*	*U2AF2*	***UBR5***	*WT1*	*XBP1*	*XPO1*	*ZNF471*	***ZRSR2***

Gene with red bold font indicates that only frameshift, stop gain/loss and inframe insertion/deletion variants are considered for these genes.

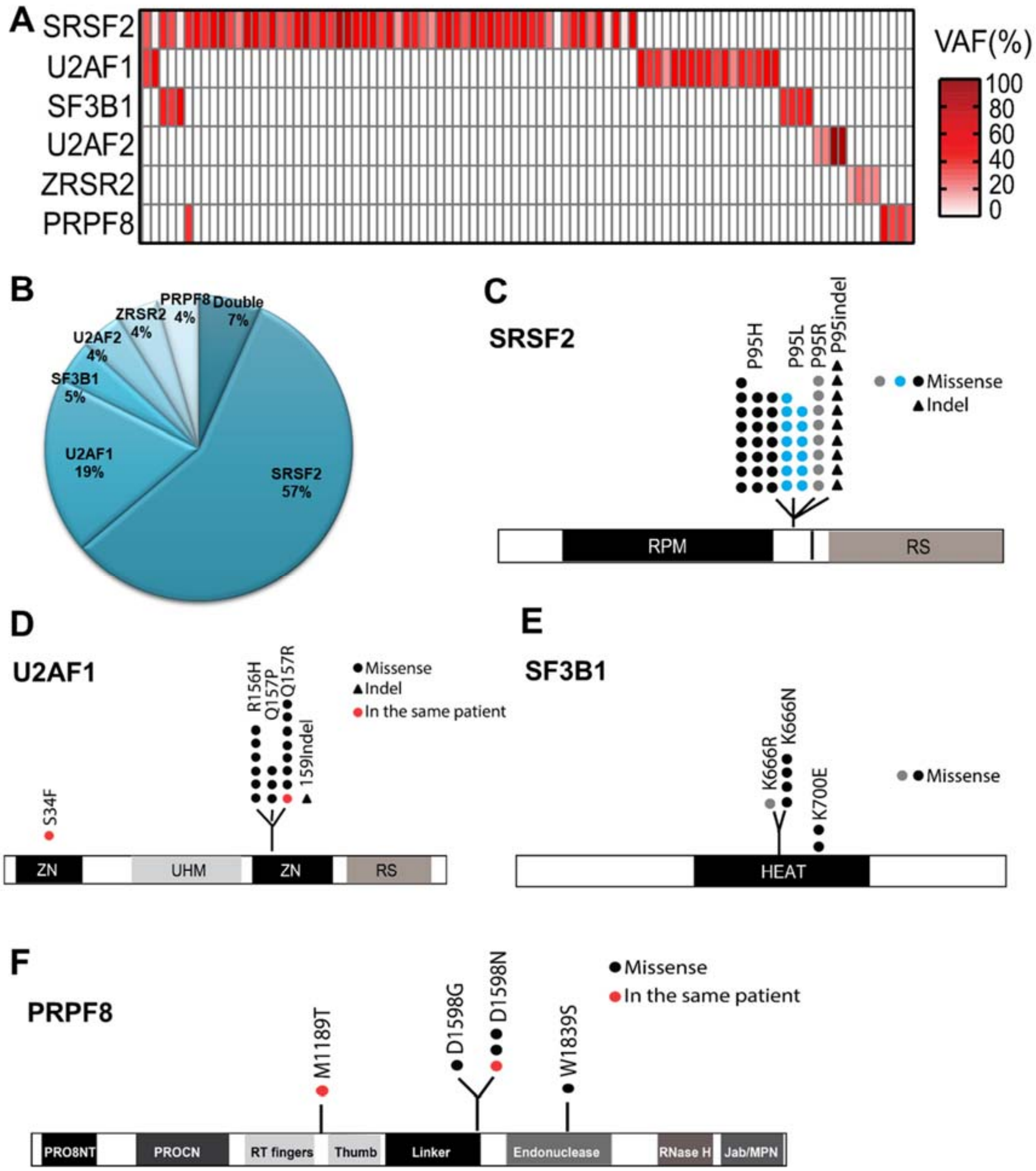
Supplementary Table 2. The frequency of gene mutations in different diagnosis groups

	CNL	aCML	Unclassifiable	CMML	CNL	aCML	Unclassifiable	CMML
	Frequency (%)				Number			
CSF3R	64.1%	22.2%	4.0%	3.4%	25	6	1	1
NRAS	10.3%	25.9%	8.0%	37.9%	18	7	2	11
JAK2	7.7%	11.1%	8.0%	3.4%	3	3	2	1
CBL	5.1%	11.1%	8.0%	17.2%	2	3	2	5
CBLB	0.0%	0.0%	0.0%	0.0%	0	0	0	0
PTPN11	10.3%	0.0%	4.0%	3.4%	4	0	1	1
KRAS	0.0%	3.7%	4.0%	10.3%	0	1	1	3
NF1	0.0%	0.0%	4.0%	3.4%	0	0	1	1
FLT3	0.0%	7.4%	4.0%	0.0%	0	2	1	0
STAT5B	0.0%	3.7%	4.0%	0.0%	0	1	1	0
ABL1	5.1%	3.7%	0.0%	0.0%	2	1	0	0
GNB1	2.6%	3.7%	0.0%	0.0%	1	1	0	0
SH2B3	0.0%	3.7%	0.0%	0.0%	0	1	0	0
JAK1	0.0%	0.0%	4.0%	3.4%	0	0	1	1
KIT	0.0%	0.0%	4.0%	0.0%	0	0	1	0
FLT3 fusion	0.0%	3.7%	0.0%	0.0%	0	1	0	0
ABL1 fusion	0.0%	0.0%	0.0%	0.0%	0	0	0	0
CALR	0.0%	0.0%	4.0%	0.0%	0	0	1	0
NTRK2	0.0%	0.0%	0.0%	0.0%	0	0	0	0
STAT3	0.0%	0.0%	0.0%	3.4%	0	0	0	1
STAT5A	0.0%	0.0%	0.0%	3.4%	0	0	0	1
CCND2	0.0%	0.0%	12.0%	0.0%	0	0	3	0
ETNK1	2.6%	3.7%	4.0%	0.0%	1	1	1	0
ASXL1	76.9%	81.5%	64.0%	69.0%	30	22	16	20
ASXL2	2.6%	3.7%	8.0%	0.0%	1	1	2	0
SRSF2	43.6%	37.0%	48.0%	24.1%	17	10	12	7
U2AF1	15.4%	14.8%	8.0%	24.1%	6	4	2	7
SF3B1	2.6%	0.0%	16.0%	0.0%	1	0	4	0
U2AF2	5.1%	0.0%	0.0%	3.4%	2	0	0	1
ZRSR2	2.6%	3.7%	0.0%	3.4%	1	1	0	1
PRPF8	2.6%	0.0%	0.0%	10.3%	1	0	0	3
TET2	20.5%	37.0%	44.0%	48.3%	8	10	11	14
SETBP1	41.0%	7.4%	16.0%	13.8%	16	2	4	4
EZH2	20.5%	29.6%	24.0%	6.9%	8	8	6	2
GATA2	12.8%	14.8%	16.0%	13.8%	5	4	4	4
RUNX1	2.6%	11.1%	4.0%	27.6%	1	3	1	8
DNMT3A	5.1%	7.4%	0.0%	10.3%	2	2	0	3
STAG2	2.6%	14.8%	8.0%	3.4%	1	4	2	1
SMC1A	5.1%	0.0%	0.0%	3.4%	2	0	0	1
RAD21	0.0%	0.0%	4.0%	0.0%	0	0	1	0
PDS5B	0.0%	3.7%	0.0%	0.0%	0	1	0	0
CUX1	5.1%	11.1%	0.0%	0.0%	2	3	0	0
PPM1D	2.6%	3.7%	0.0%	0.0%	1	1	0	0
TP53	2.6%	0.0%	0.0%	0.0%	1	0	0	0
BRCC3	2.6%	0.0%	4.0%	0.0%	1	0	1	0
NPM1	0.0%	3.7%	0.0%	3.4%	0	1	0	1
CEBPA	0.0%	0.0%	8.0%	0.0%	0	0	2	0
IDH2	2.6%	0.0%	0.0%	0.0%	1	0	0	0
NFE2	0.0%	0.0%	0.0%	0.0%	0	0	0	0
WT1	5.1%	0.0%	0.0%	0.0%	2	0	0	0
PHF6	2.6%	3.7%	0.0%	6.9%	1	1	0	2
BCOR	0.0%	3.7%	4.0%	3.4%	0	1	1	1
BCORL1	0.0%	0.0%	0.0%	3.4%	0	0	0	1
Total number					39	27	25	29

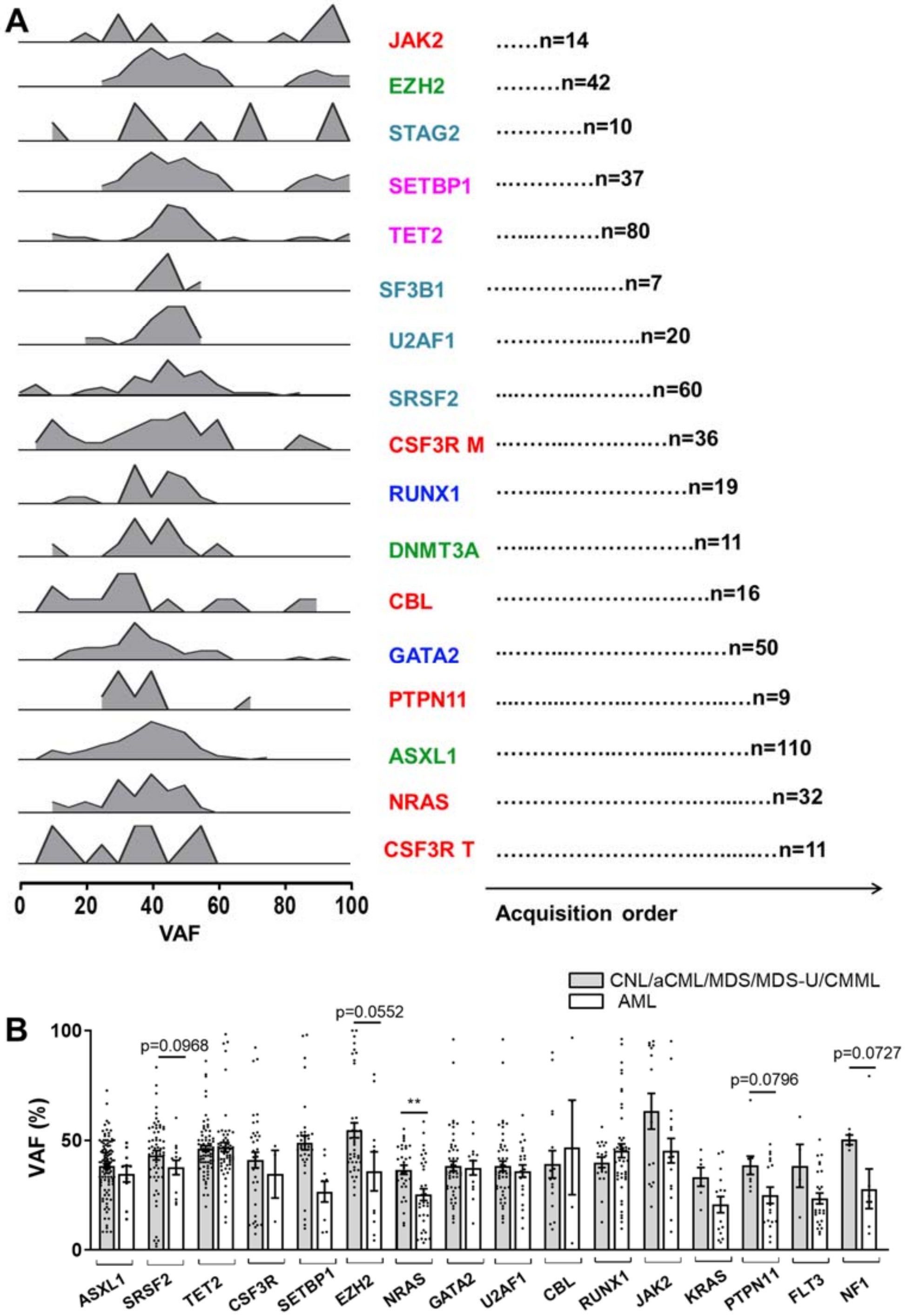
*represents statistical significance. Statistical analysis was performed using contingency table Chi-Square and Bonferroni multiple comparison correction.



Supplementary Figure 1. Sanger validation of *ASXL1* G643GX mutation. The graph depicts the Sanger sequencing validation of *ASXL1* G643GX mutations detected by exome sequencing (bottom seven samples). The top three samples are control samples (no ASXL1 mutations detected by exome sequencing). Variant allelic frequencies detected by exome sequencing were shown.



Supplementary Figure 2. Distribution of mutations on splicing factors.
(A) The mosaic plot depicts the spectrum of different splicing factor in the cohorts. (B) The pie chart depicts the frequencies of different splicing factor mutations. The graph depicts the structure and distributions of mutations on *SRSF2* (C), *U2AF1* (D), *SF3B1* (E), and *PRPF8* (F).



Supplementary Figure 3. Clonal architecture of different pathway mutations. (A) The histogram illustrates VAF and the number of patients with a particular gene mutation. Gene mutations with higher VAFs are considered to occur earlier then variants with lower VAFs. (B) The graph depicts Mean ± SEM of VAFs of common driver mutations in CNL/aCML/unclassifiable/CMML from the current study and AML patients from the BeatAML study. Statistical analysis was performed using two-tailed Mann-Whitney tests and expressed as ** p<.01.

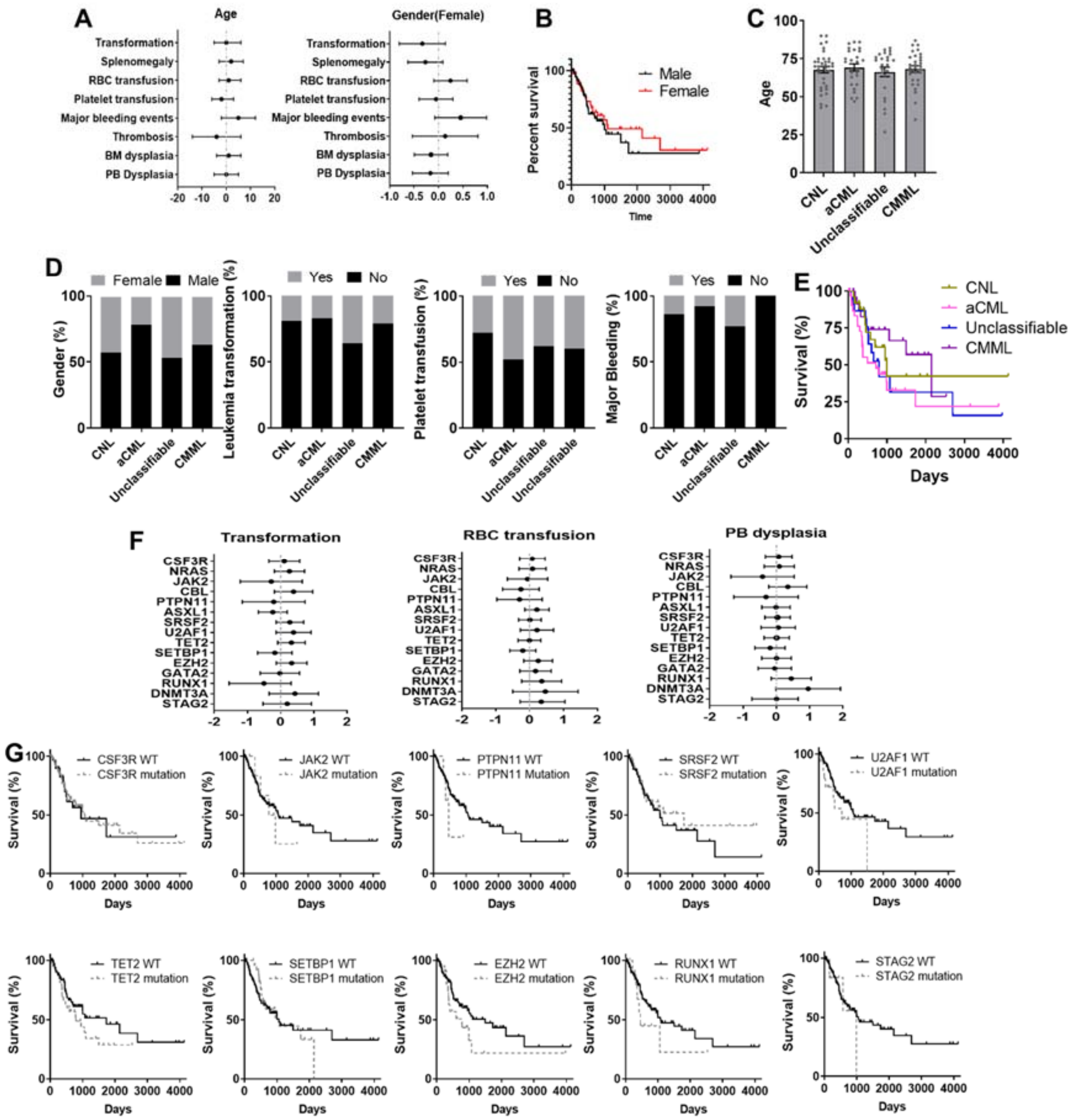
Supplementary Table 3. The frequency of gene mutations in different signaling molecular groups

	RAS	***CSF3R***	***JAK2***	***RAS***	***CSF3R***	***JAK2***
		Frequency			number	
CNL	12.5%	77.8%	25.0%	6	21	3
aCML	14.6%	11.1%	25.0%	7	3	3
Unclassifiable	14.6%	3.7%	16.7%	7	1	2
CMML	39.6%	3.7%	8.3%	19	1	1
ASXL1	72.9%	85.2%	50.0%	35	23	6
ASXL2	4.2%	3.7%	0.0%	2	1	0
SRSF2	45.8%	40.7%	16.7%	22	11	2
U2AF1	12.5%	18.5%	0.0%	6	5	0
SF3B1	0.0%	3.7%	8.3%	0	1	1
U2AF2	0.0%	11.1%	0.0%	0	3	0
ZRSR2	0.0%	7.4%	16.7%	0	2	2
TET2	41.7%	22.2%	50.0%	20	6	6
SETBP1	20.8%	40.7%	0.0%	10	11	0
EZH2	18.8%	14.8%	33.3%	9	4	4
GATA2	20.8%	11.1%	0.0%	10	3	0
RUNX1	18.8%	7.4%	8.3%	9	2	1
DNMT3A	8.3%	7.4%	8.3%	4	2	1
STAG2	12.5%	3.7%	0.0%	6	1	0
PPM1D	0.0%	0.0%	16.7%	0	0	2
WT1	0.0%	7.4%	0.0%	0	2	0
NPM1	4.2%	0.0%	0.0%	2	0	0
SMC1A	0.0%	0.0%	0.0%	0	0	0
PRPF8	2.1%	3.7%	0.0%	1	1	0
PDS5B	0.0%	0.0%	0.0%	0	0	0
RAD21	0.0%	0.0%	8.3%	0	0	1
CUX1	2.1%	3.7%	8.3%	1	1	1
TP53	2.1%	0.0%	0.0%	1	0	0
BRCC3	0.0%	3.7%	0.0%	0	1	0
CEBPA	0.0%	0.0%	0.0%	0	0	0
IDH2	0.0%	0.0%	0.0%	0	0	0
NFE2	0.0%	0.0%	0.0%	0	0	0
PHF6	4.2%	0.0%	8.3%	2	0	1
BCOR	4.2%	0.0%	8.3%	2	0	1
BCORL1	0.0%	0.0%	8.3%	0	0	1
Total number				48	27	12

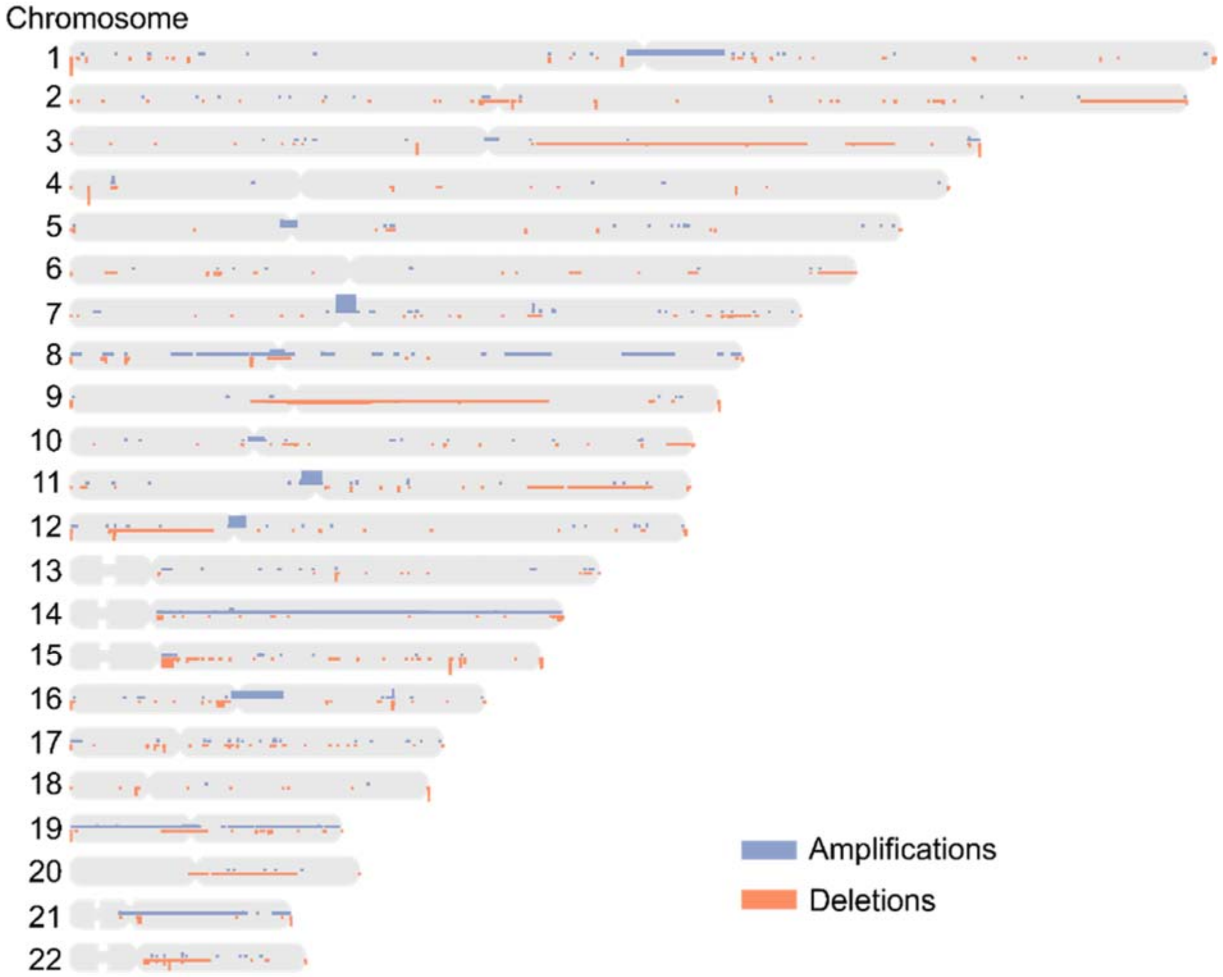
Supplementary Table 4. Coexisting different signaling pathway mutations in CNL/aCML/unclassifiable/CMML

	ID	Signaling gene_1	Signaling gene_2	Signaling gene_3	Other gene_1	Other gene_2	Other gene_3	Other gene_4	Other gene_5	Other gene_6
1	13-00256	NF1	JAK2	SH2B3*	TET2	BCOR				
2	13-00187	CSF3R	NRAS	NTRK2	ASXL1	SETBP1				
3	15-00467	CSF3R*	ABL1		ASXL1	DNMT3A	ASXL2			
4	14-00201	CSF3R*	CALR		ASXL1	SRSF2	TET2	WT1		
5	14-00804	CSF3R	CBL		ASXL1	U2AF1	DNMT3A			
6	13-00514	CSF3R	CBL		ASXL1					
7	08-00423	CSF3R*	CBL		ASXL1	U2AF1				
8	13-00037	CSF3R	NRAS		ASXL1	SRSF2	EZH2	TET2		
9	12-00364	CSF3R	NRAS		ASXL1	EZH2	SETBP1	TET2		
10	13-00369	CSF3R_T	NRAS		ASXL1	SRSF2	SETBP1			
11	13-00438	CSF3R	PTPN11		ASXL1	SETBP1	EZH2			
12	12-00212	CSF3R	PTPN11		ASXL1	SETBP1	EZH2*			
13	14-00389	CSF3R	PTPN11		ASXL1	SRSF2	SETBP1	TET2		
14	15-00270	NRAS	ABL1		ASXL1	U2AF1	GATA2			
15	14-00413	NRAS	CBL		ASXL1	SRSF2	SETBP1	TET2		
16	14-00131	NRAS	CBL		ASXL1	TET2*	EZH2*	STAG2		
17	14-00685	NRAS	FLT3		NPM1	DNMT3A				
18	12-00388	NRAS	GNB1		ASXL2*	SRSF2				
19	13-00359	NRAS	KRAS		ASXL1	EZH2*	TET2	RUNX1		
20	12-00370	NRAS	KRAS		ASXL1	SRSF2	TET2			
21	13-00023	NRAS	STAT3		ASXL1	TET2				
22	09-00020	NRAS	STAT5A		ASXL1	SRSF2	U2AF1	GATA2*	RUNX1	DNMT3A
23	14-00516	KRAS	JAK1		ASXL1	U2AF1*	GATA2*			
24	13-00269	PTPN11	NF1		ASXL1	SRSF2	SETBP1	CUX1		

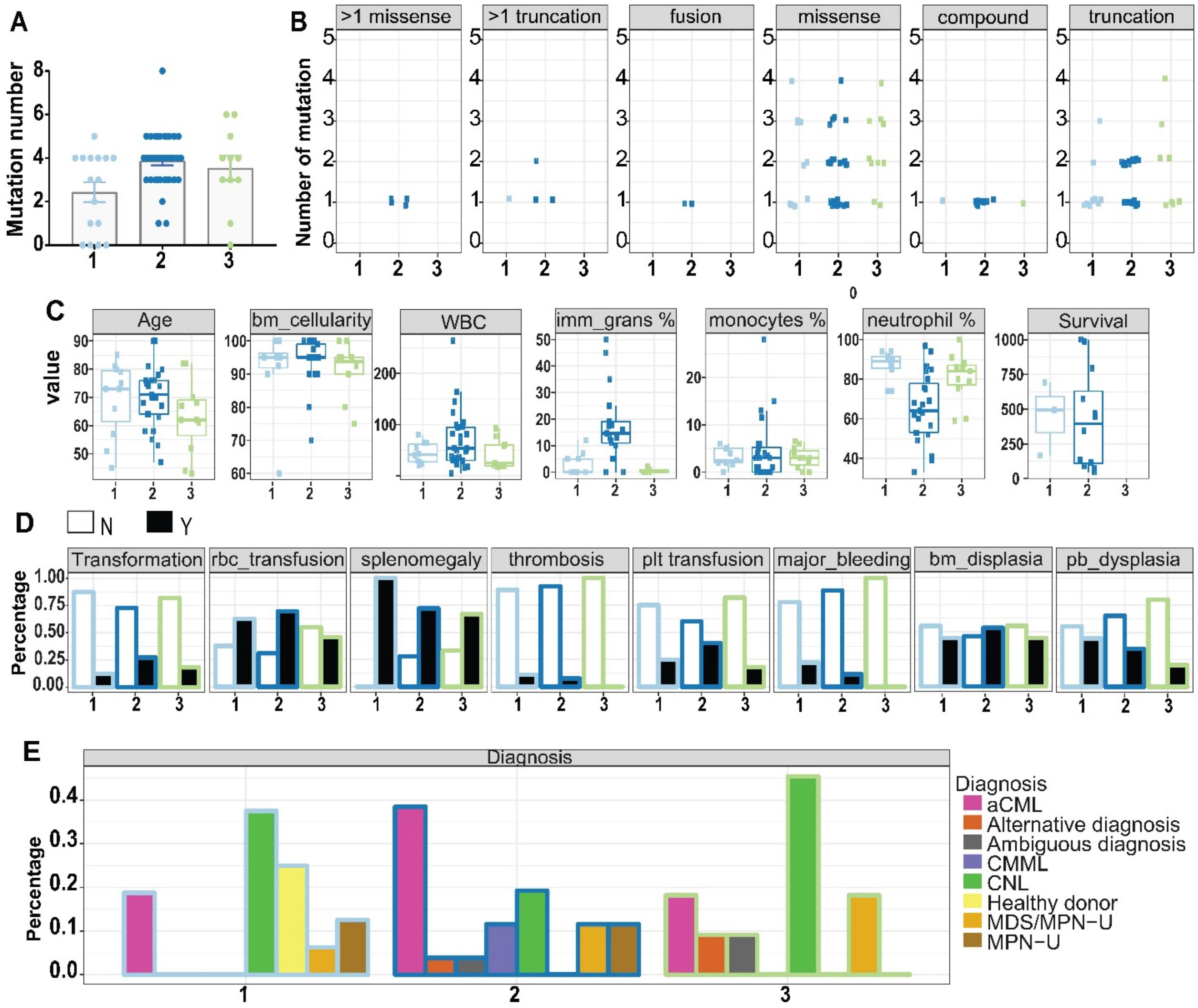
* indicates the presence of more than one mutation. *CSF3R_T*: *CSF3R* truncation mutation.



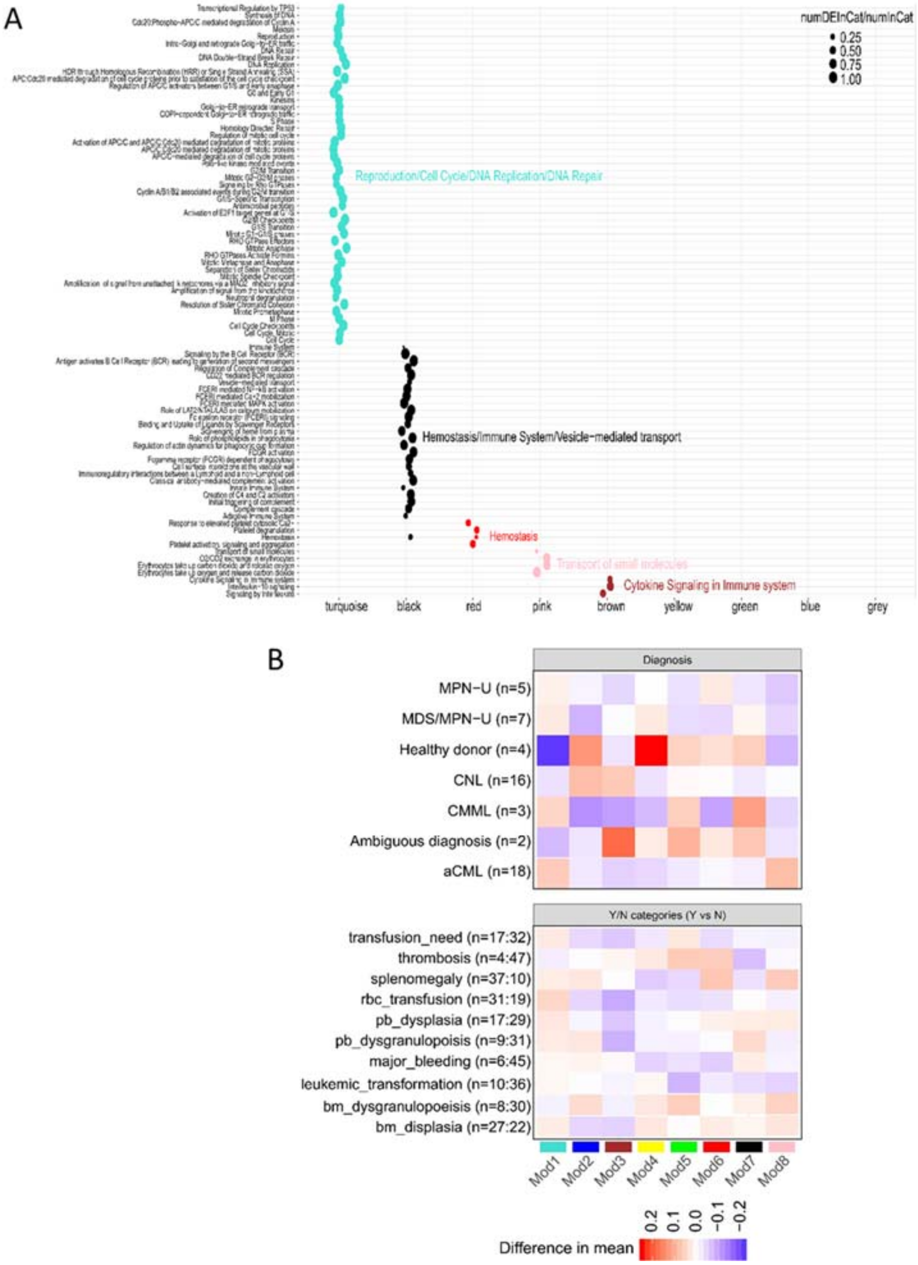
Supplementary Figure 4. The association between clinical parameters and clinical outcomes. (A) Graphs depict 95% CI and Hodges-Lehmann median difference of odds ratios of age and gender for different clinical outcomes calculated by Fisher's exact tests. (B) The graph depicts the Kaplan-Meier survival curve of patients with male or female gender. Statistical significance was analyzed by the log-rank test. (C) Graphs depict the mean ± SEM of age in different disease subgroups. Statistical significance was assessed using one-way ANOVA and Kruskal-Wallis tests. (D) Graphs depict the comparison of frequencies of indicated clinical outcomes in different disease groups. Statistical significance was analyzed using a contingency table chi-square test. (E) The graph depicts the Kaplan-Meier survival curve of patients in different diagnosis subgroups. Statistical significance was analyzed by a log-rank test. (F) The graph depicts 95% CI and the median difference of the log-transformed odds ratios for different clinical parameters in the presence or absence of mutations in a given gene calculated by Fisher's exact tests. (G) Graphs depict the Kaplan-Meier survival curve of patients in the presence or absence of given gene mutations. Statistical significance was analyzed by log-rank tests.



Supplementary Figure 5. CNV analysis. The plot displays CNV regions determined by log2 (tumor read count/pooled normal count) on human chromosome 1-22, keeping only those with an absolute value >0.5. The height of the colored bar corresponds to the count of samples with a CNV in that region, with a maximum of 32. Segments in orange represent deletions, whereas purple segments indicate duplications. Grey regions represent normal regions, where there was not enough evidence to call a copy number variant. For more detailed information, please visit our online interactive user interface, Vizome, at [www.vizome.org].



Supplementary Figure 6. Differential clinical parameters in different consensus clusters. (A) The graph plotted the number of mutated genes in each of the cluster. (B) Frequency of mutation classes by the cluster. For each mutation class, the number of genes in that class per sample (Y-axis) is shown relative to the cluster membership of the sample (X-axis and color). (C) Boxplots of the distribution of the numeric clinical data relative to the Consensus Clustering (k=7) clusters. Each data point indicates the value of the indicated clinical variable per sample. (D) The proportion of whether or not a given categorical clinical value (separated by subplots) was considered yes (Y) or no (N) in the 3 largest Consensus Cluster groups. (E) The proportion of the curated diagnosis categories with respect to the 3 largest clusters. The bars are filled with the diagnosis colors with the outlines indicating the cluster colors.



Supplementary Figure 7. Reactome pathway and WGCNA gene expression and clinic parameter analysis. (A) Reactome pathway analysis for the WGCNA modules. Significant Reactome pathways for the WGCNA modules at a Benjamini-Hochberg (BH) FDR < .05. The size of the points indicates the proportion of pathway genes that are also in a given module. The text to the right of each module indicates the highest level Reactome pathway(s) significantly enriched in the module. (B) Heatmap summary of the eigengene differences between the clinical categorical variables. The categories are grouped by either ‘diagnosis' or ‘Y/N' categories. The diagnosis categories indicate the difference between the average of the eigengenes of the given category for each module vs the average of the remaining categories. The sample size is indicated in parentheses. The ‘Y/N' categories indicate the average difference between the ‘Y' groups vs the ‘N' group in terms of the module eigengenes. The sample size is shown as (Y: N). The x-axis indicates module name and color.

Supplementary References

1. Li H. Aligning sequence reads , clone sequences and assembly contigs with BWA-MEM. 2013;
2. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–1303.
3. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865–2871.
4. Zhang H, Savage S, Schultz AR, et al. Clinical resistance to crenolanib in acute myeloid leukemia due to diverse molecular mechanisms. *Nat. Commun.* 2019;10(1):244.
5. Jaiswal S, Fontanillas P, Flannick J, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* 2014;371(26):2488–2498.
6. Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–74.
7. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018;562(7728):526–531.
8. Zhang H, Reister Schultz A, Luty S, et al. Characterization of the leukemogenic potential of distal cytoplasmic CSF3R truncation and missense mutations. *Leukemia*. 2017;
9. Liao Y, Smyth GK, Shi W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* 2013;41(10):.
10. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Comput. Biol.* 2016;12(4):e1004873.
11. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. 2007;23(6):657–663.
12. Haas B, Dobin A, Stransky N, et al. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv*. 2017;120295.
13. Kim D, Salzberg SL. TopHat-Fusion: An algorithm for discovery of novel fusion transcripts. *Genome Biol.* 2011;12(8):.
14. Zhang H, Paliga A, Hobbs E, et al. Two myeloid leukemia cases with rare FLT3 fusions. *Cold Spring Harb. Mol. case Stud.* 2018;4(6):.
15. Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* 2003;52(1–2):91–118.
16. Law CW, Chen Y, Shi W, Smyth GK. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014;15(2):.
17. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010;11(2):.
18. Fabregat A, Sidiropoulos K, Viteri G, et al. Reactome pathway analysis: A high-performance in-memory approach. *BMC Bioinformatics*. 2017;18(1):.